Published online 20 October 2009 Nucleic Acids Research, 2010, Vol. 38, Database issue D457–D462
doi:10.1093/nar/gkp851

TriTrypDB: a functional genomic resource for the Trypanosomatidae

Martin Aslett[1], Cristina Aurrecoechea[2], Matthew Berriman[1], John Brestelli[3], Brian P. Brunk[3], Mark Carrington[4], Daniel P. Depledge[1], Steve Fischer[3], Bindu Gajria[3], Xin Gao[3], Malcolm J. Gardner[5,6], Alan Gingle[7], Greg Grant[3], Omar S. Harb[3,*], Mark Heiges[2], Christiane Hertz-Fowler[1,*], Robin Houston[1], Frank Innamorato[3], John Iodice[3], Jessica C. Kissinger[2,8], Eileen Kraemer[9], Wei Li[3], Flora J. Logan[1], John A. Miller[9], Siddhartha Mitra[5], Peter J. Myler[5,6,10], Vishal Nayak[3], Cary Pennington[2], Isabelle Phan[5], Deborah F. Pinney[3], Gowthaman Ramasamy[5], Matthew B. Rogers[1], David S. Roos[11], Chris Ross[2], Dhileep Sivam[5], Deborah F. Smith[12], Ganesh Srinivasamoorthy[2], Christian J. Stoeckert Jr[3], Sandhya Subramanian[5], Ryan Thibodeau[2], Adrian Tivey[1], Charles Treatman[3], Giles Velarde[1] and Haiming Wang[2]

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK, [2]Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, GA 30602, [3]Penn Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA 19104 USA, [4]Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK, [5]Seattle Biomedical Research Institute, Seattle, WA 98109, [6]Department of Global Health, University of Washington, Seattle, WA 98195, [7]Center for Applied Genetic Technologies, [8]Department of Genetics, [9]Department of Computer Science, University of Georgia, Athens, GA 30602, [10]Department of Medical Education and Biomedical Informatics, University of Washington, Seattle, WA 98195, [11]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104 USA and [12]Centre for Immunology and Infection, Department of Biology, University of York, Heslington, York YO10 5YW, UK

Received August 15, 2009; Accepted September 23, 2009

ABSTRACT

TriTrypDB (http://tritrypdb.org) is an integrated database providing access to genome-scale datasets for kinetoplastid parasites, and supporting a variety of complex queries driven by research and development needs. TriTrypDB is a collaborative project, utilizing the GUS/WDK computational infrastructure developed by the Eukaryotic Pathogen Bioinformatics Resource Center (EuPathDB.org) to integrate genome annotation and analyses from GeneDB and elsewhere with a wide variety of functional genomics datasets made available by members of the global research community, often pre-publication. Currently, TriTrypDB integrates datasets from *Leishmania braziliensis, L. infantum, L. major, L. tarentolae, Trypanosoma brucei* and *T. cruzi*. Users may examine individual genes or chromosomal spans in their genomic context, including syntenic alignments with other kinetoplastid organisms. Data within TriTrypDB can be interrogated utilizing a sophisticated search strategy system that enables a user to construct complex queries combining multiple data types. All search strategies are stored, allowing future access and integrated searches. 'User Comments' may be added to any gene page, enhancing available annotation; such comments become immediately searchable via the text search, and are forwarded to curators for incorporation into the reference annotation when appropriate.

INTRODUCTION

The *Trypanosomatidae* are a group of unicellular, flagellated, obligate parasites, including many important pathogens of humans and animals. African trypanosomes (*Trypanosoma brucei, T. congolense* and *T. vivax*) are endemic in rural areas of sub-Saharan Africa, where they cause sleeping sickness in humans, and wasting

*To whom correspondence should be addressed. Tel: +1 215 746 7019; Fax: +1 215 573 3111; Email: oharb@pcbi.upenn.edu
Correspondence may also be addressed to Christiane Hertz-Fowler. Tel: +44 (0)1223 834244; Fax: +44 (0)1223 494919; Email: chf@sanger.ac.uk

disease (nagana) in cattle. These diseases are invariably fatal if left untreated (1). In 2004, 17 580 cases of human infection were reported, but due to chronic under-reporting in poor, rural areas, actual infection rates are thought to reach 300 000 new cases annually (2,3). Millions of cattle are also at risk, and trypanosomiasis severely constrains cattle grazing in endemic regions. *Trypanosoma cruzi* is endemic in south and central America, causing Chagas disease (American Trypanosomiasis) in approximately 8–9 million infected individuals, and ~14 000 deaths annually (4). *Leishmania* parasites are found throughout the world (old world: *L. major*; new world: *L. infantum* and *L. braziliensis*), infecting an estimated 12 million individuals, with approximately 2 million new cases reported annually (5). These parasites exhibit a variety of spectral pathologies, including severely debilitating cutaneous disease, and visceral symptoms that may be fatal.

The genomes of multiple Trypanosomatidae have been sequenced (6–9) and are available from sources such as GeneDB (http://GeneDB.org) (10) and the primary sequence nucleotide databases (DDBJ, EMBL and GenBank); genome projects are also underway for a variety of other species of biological and evolutionary interest. Whilst GeneDB specializes in the display of highly-curated annotation, it has been difficult to integrate this information with other available 'omics' datasets: expression profiling data, proteomics results, etc., hindering the research on these organisms, including the development of new therapies and diagnostics. To this end, the TriTrypDB initiative was undertaken as a collaborative effort between the EuPathDB team at the Universities of Pennsylvania and Georgia (http://EuPathDB.org) (11), the GeneDB group at the Wellcome Trust Sanger Institute and researchers from the Seattle Biomedical Research Institute, culminating in the release of the first version of TriTrypDB (http://TriTrypDB.org) in early 2009. This collaboration has proved to be an effective means for providing the scientific community with up-to-date annotation and curation, and access to tools enabling sophisticated queries against genomic scale datasets.

DATA IN THE CURRENT RELEASE

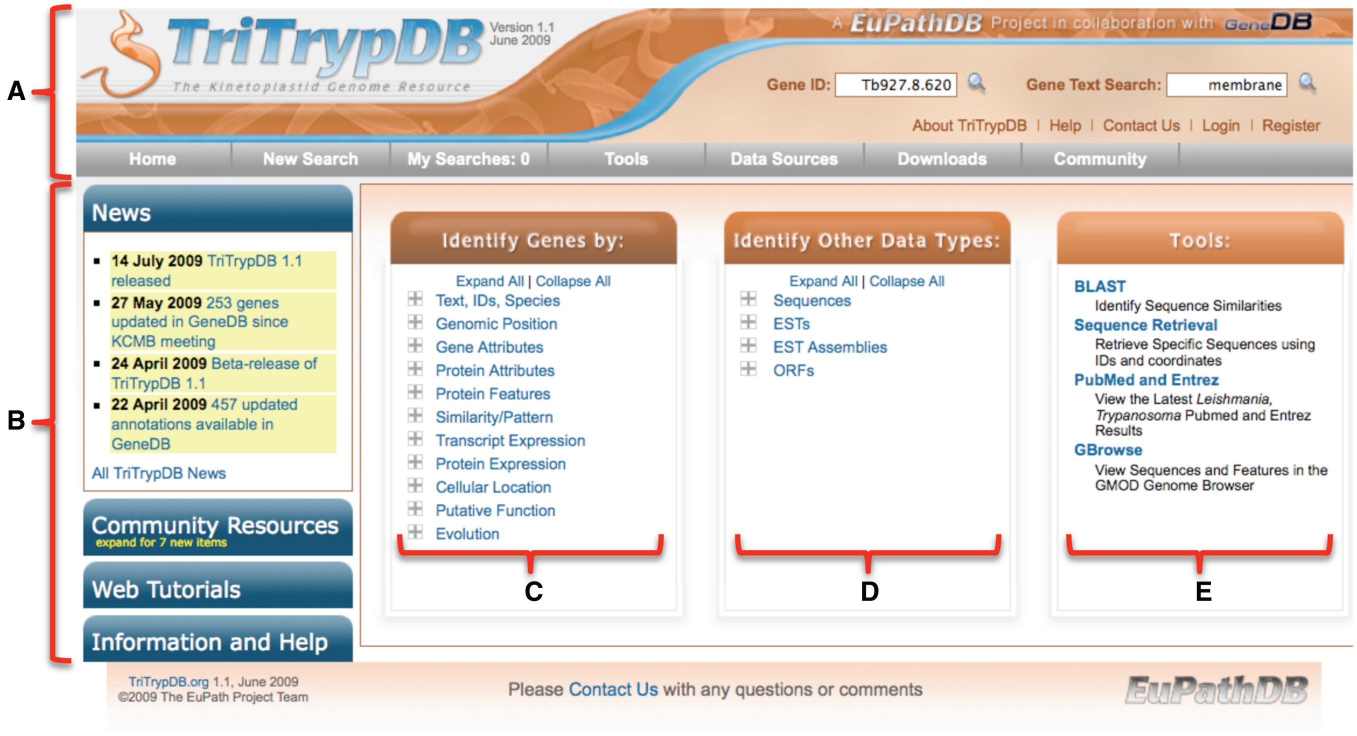
TriTrypDB (release 1.1) houses the genome sequences of *T. brucei* TREU 927 strain [11 chromosomes, 26 megabases (Mb)] (6); *T. cruzi* CL Brener, Esmeraldo and non-Esmeraldo-like haplotypes (41 chromosomes, 67 Mb) (7,12); *L. major* Friedlin strain (36 chromosomes, 32.8 Mb) (8), *L. infantum* JPCM5 strain (36 chromosomes, 32 Mbs) (9); *L. braziliensis* Viannia strain (35 chromosomes, 32 Mb) (9); and *L. tarentolae* (sequence kindly provided in advance of publication, by Marc Ouellette, Université Laval, and Martin Olivier from McGill University). TriTrypDB also includes selected transcript and proteomics expression data (with more to follow over the coming months). Transcript expression information is derived from both microarray data (*L. infantum* differentiation time series and data provided pre-publication by Dan Zilberstein and Peter Myler) and expressed sequence tags (EST libraries from *T. brucei*, *T. cruzi*, *L. braziliensis*, *L. infantum* and *L. major* extracted from dbEST; http://www.ncbi.nlm.nih.gov/dbEST). Protein expression data based on tandem mass spectrometry of whole parasites and subcellular fractions is available for *T. brucei* (13), *T. cruzi* (14), *L. braziliensis*, *L. infantum* and *L. major* [(15), and Marc Ouellette, pre-publication].

USING TriTrypDB

The home page of TriTrypDB is based on the recently re-designed EuPathDB web page, and includes five main sections (Figure 1). The top of the page is an interactive banner (Figure 1A), which appears on all pages and includes (i) the TriTrypDB logo, (ii) windows for ID and text searches, (iii) links providing quick access to useful pages (help and information pages, a 'Contact Us' link and links for registration/login) and (iv) a tool bar (grey) with links to access diverse searches (see below), the user's personal search history, tools, downloads, data sources and other links. The left side of the home page (Figure 1B) provides a series of expandable windows presenting news items (such as release notes), tutorials (demonstrating website usage), community resources and additional information and help. New items added since the user's last visits are indicated by yellow numbers. Three panels in the middle of the page provide access to searches and tools: the panel indicated as Figure 1C includes diverse searches pertaining to genes; Figure 1D accesses searches against other data types (assemblies, ESTs, Open Reading Frame (ORFs), etc., with more data types to follow, as already implemented for other EuPathDB component databases); Figure 1E provides links to tools such as BLAST, sequence retrieval and a genome browser.

Visitors to TriTrypDB may select from approximately 80 different searches against the TriTryp genomes and datasets. Importantly, searches can be combined in an integrated and graphical manner (Figure 2A and B), and results are displayed in tabular lists below the growing search strategy (Figure 2C and D). In the example presented as Figure 2, the search strategy begins with a text search for the word kinase using either the 'Gene Text Search' window in the interactive banner (Figure 1A), or accessing the text search query page from 'Identify Genes by' section (Figure 1C). The latter provides a greater range of user options, such as defining which fields to search. All searches are also accessible by positioning the (mouse) cursor over 'New Search' in the banner tool bar (Figure 1A); for example, users seeking kinases may also wish to consider searching GO annotations, Interpro domains, etc. The text query presented in Figure 2 yields 1986 gene records, in any of the species supported by TriTrypDB, which contain the word kinase, and is displayed as a graphical image in the search strategy window (step 1 of Figure 2A).

The user may be satisfied with this result, or may wish to revise or combine it with other searches: for example,



Gene ID: Tb927.8.620

Gene Text Search: membrane

- **14 July 2009** TriTrypDB 1.1 released
- **27 May 2009** 253 genes updated in GeneDB since KCMB meeting
- **24 April 2009** Beta-release of TriTrypDB 1.1
- **22 April 2009** 457 updated annotations available in GeneDB

BLAST: Identify Sequence Similarities

Sequence Retrieval: Retrieve Specific Sequences using IDs and coordinates

PubMed and Entrez: View the Latest *Leishmania*, *Trypanosoma* Pubmed and Entrez Results

GBrowse: View Sequences and Features in the GMOD Genome Browser

TriTrypDB.org 1.1, June 2009
©2009 The EuPath Project Team

Please Contact Us with any questions or comments

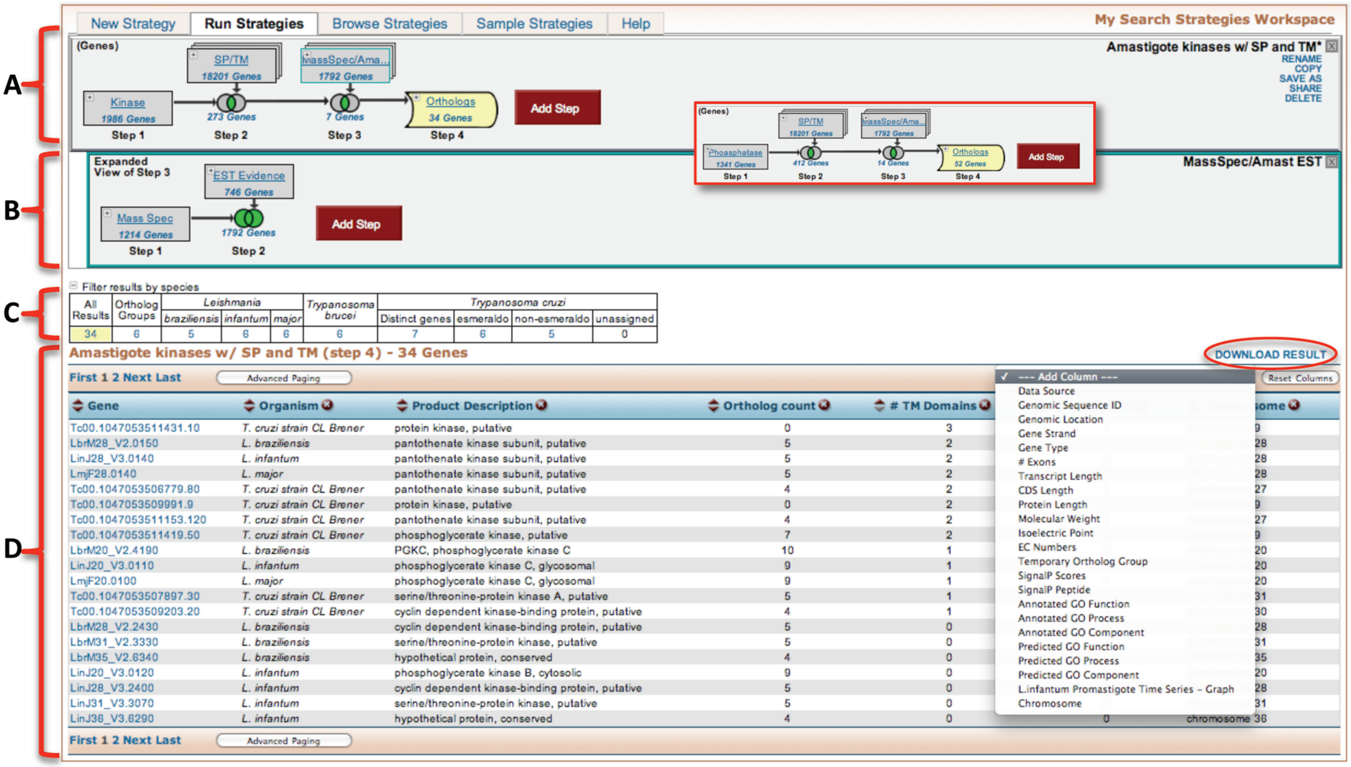
Figure 1. Screen shot of the TriTrypDB home page. (**A**) Interactive banner present on all TriTrypDB web pages, including quick search windows and a tool bar (grey). (**B**) Side bar components contain expandable sections for release notes, community resources, tutorials and help (new items are highlighted with a yellow alert). (**C**) Gene searches; clicking on '+' symbols reveals a list of searches available within each category. (**D**) Searches of non-gene entities, such as ESTs and ORFs of genome sequence. (**E**) Links to available tools, including the genome browser (based on GBrowse), BLAST against TriTrypDB, the sequence retrieval tool and recent PubMed records pertaining to TriTrypDB organisms.

how many of these kinases are predicted to be secreted? Clicking on the 'add step' button opens a pop-up window containing all available searches, from which the user can select the 'cellular location' query 'predicted signal peptide' (data not shown), specify appropriate parameters (or accept the defaults) and select how to combine this search with the previous one in the strategy (i.e. which Boolean operator to use: intersection, union or minus). Intersecting genes predicted to contain a signal peptide with those containing the word kinase are displayed, along with a Venn diagram representing how these searches were combined, in step 2 (Figure 2A). This strategy can be further expanded by asking which of these results is supported by proteomics evidence in general or by proteomics evidence from specific parasite life cycle stages as shown in Figure 2 (Step 3).

This search yields a limited number of hits, for several reasons, including the limited amount of functional genomics evidence available. The search can readily be expanded, however, as indicated in the main body of Figure 2. For example, the signal peptide search can be expanded to consider genes that contain a signal peptide and/or a transmembrane domain (as it is not clear that all secreted proteins are properly annotated and accurately recognized by SignalP). Similarly, expression data can be expanded by considering genes with either proteomics evidence or EST support (this information appears as a nested 'sub-strategy' 3 in Figure 2B). The ability to apply an ortholog transform on any search result (i.e. identify orthologs for a set of genes), based on orthologs identified by OrthoMCL (http://orthomcl.org) (16), provides another powerful method for expanding searches. In the strategy shown, an ortholog transform identifies any secreted kinetoplastid parasite gene for which expression evidence is available for an ortholog in any other member of the kinetoplastida (Step 3, Figure 2A).

All steps, in all searches, may be revised, renamed, transformed, deleted or expanded as nested sub-strategies. Nested strategies allow a user to expand a specific step as a separate branch of the strategy. For example, revising the first step in Figure 1 to substitute phosphatase for kinase results in changes which are propagated through all subsequent steps (red inset).

The results of a search strategy are summarized by species (filter table) (Figure 2C), and displayed as an interactive gene list (Figure 2D). The filter table provides a bird's-eye view of results of a search across all species in TriTypDB, and allows the user to click on any results in this table to display a particular species in the gene list shown below. Similarly, clicking on the results shown in any of the graphical icons in Figure 2A and B changes the table and gene list (Figure 2C and D). The gene list may also be modified, by adding, deleting or moving columns (by dragging them to the preferred position). Finally, all results can also be downloaded by clicking on the 'download results' link.



⊟ Filter results by species

All Results	Ortholog Groups	*Leishmania*			*Trypanosoma brucei*	*Trypanosoma cruzi*			
		braziliensis	*infantum*	*major*		Distinct genes	esmeraldo	non-esmeraldo	unassigned
34	6	5	6	6	6	7	6	5	0

Amastigote kinases w/ SP and TM (step 4) - 34 Genes

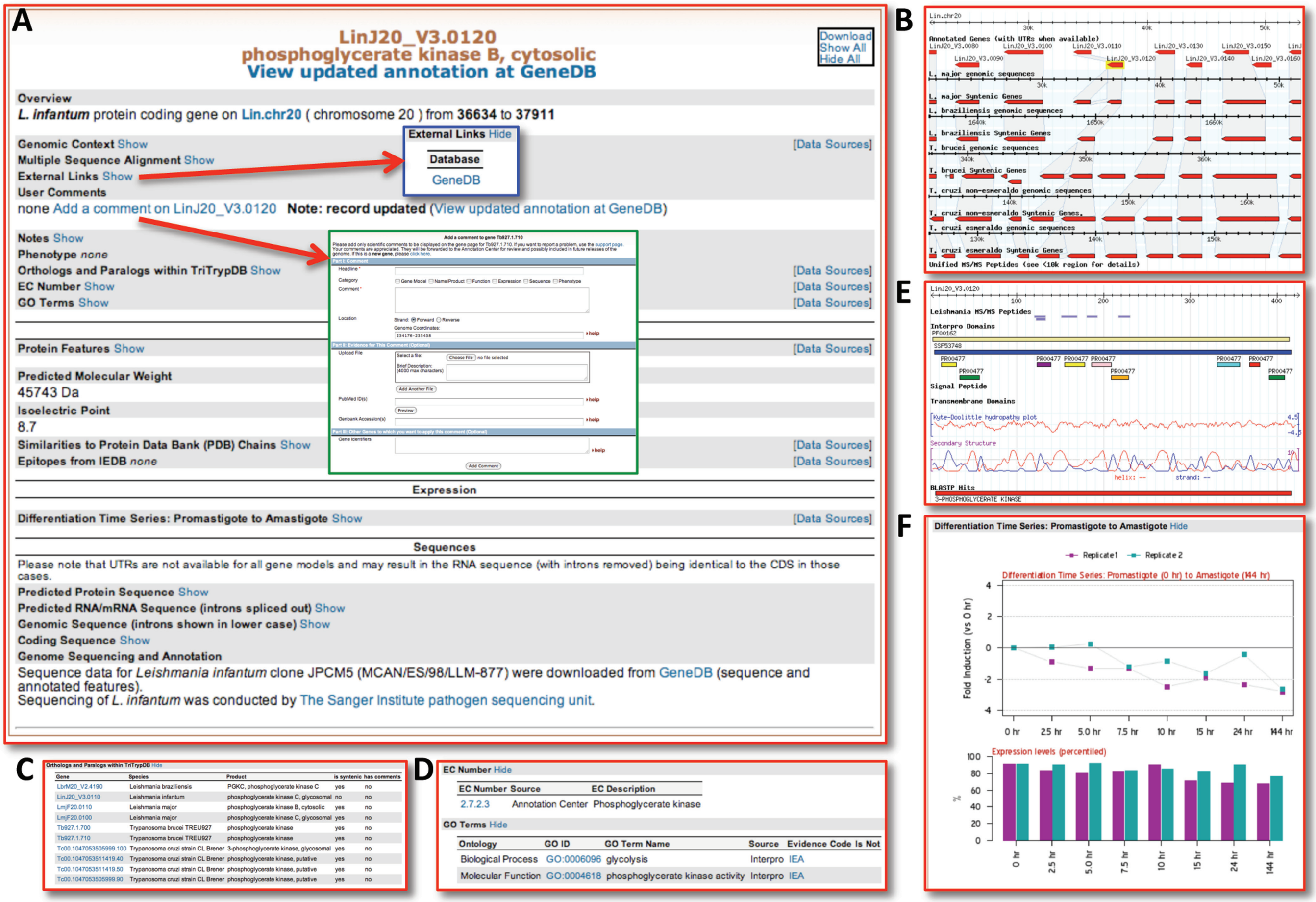
Gene	Organism	Product Description	Ortholog count	# TM Domains		ome
Tc00.1047053511431.10	*T. cruzi strain CL Brener*	protein kinase, putative	0	3		9
LbrM28_V2.0150	*L. braziliensis*	pantothenate kinase subunit, putative	5	2		28
LinJ28_V3.0140	*L. infantum*	pantothenate kinase subunit, putative	5	2		28
LmjF28.0140	*L. major*	pantothenate kinase subunit, putative	5	2		28
Tc00.1047053506779.80	*T. cruzi strain CL Brener*	pantothenate kinase subunit, putative	4	2		27
Tc00.1047053509991.9	*T. cruzi strain CL Brener*	protein kinase, putative	0	2		9
Tc00.1047053511153.120	*T. cruzi strain CL Brener*	pantothenate kinase subunit, putative	4	2		27
Tc00.1047053511419.50	*T. cruzi strain CL Brener*	phosphoglycerate kinase, putative	7	2		9
LbrM20_V2.4190	*L. braziliensis*	PGKC, phosphoglycerate kinase C	10	1		20
LinJ20_V3.0110	*L. infantum*	phosphoglycerate kinase C, glycosomal	9	1		20
LmjF20.0100	*L. major*	phosphoglycerate kinase C, glycosomal	9	1		20
Tc00.1047053507897.30	*T. cruzi strain CL Brener*	serine/threonine-protein kinase A, putative	5	1		31
Tc00.1047053509203.20	*T. cruzi strain CL Brener*	cyclin dependent kinase-binding protein, putative	4	1		30
LbrM28_V2.2430	*L. braziliensis*	cyclin dependent kinase-binding protein, putative	5	0		28
LbrM31_V2.3330	*L. braziliensis*	serine/threonine-protein kinase, putative	5	0		31
LbrM35_V2.6340	*L. braziliensis*	hypothetical protein, conserved	4	0		35
LinJ20_V3.0120	*L. infantum*	phosphoglycerate kinase B, cytosolic	9	0		20
LinJ28_V3.2400	*L. infantum*	cyclin dependent kinase-binding protein, putative	5	0		28
LinJ31_V3.3070	*L. infantum*	serine/threonine-protein kinase, putative	5	0		31
LinJ36_V3.6290	*L. infantum*	hypothetical protein, conserved	4	0	0	chromosome 36

Figure 2. Screen shot of the search strategy and results summary page. (**A**) The expanding search strategy—a search strategy is built by adding steps, which constitute a search combined with the previous step using Boolean operators (intersect, union and minus). Any step in a strategy may be revised, deleted or expanded—the insert in the red box shows the effect of revising the first step in the strategy in A. (**B**) An example of a nested strategy—the search feeding into Step 3 in (A) was expanded to include other searches without the need to re-run the entire strategy. (**C**) The filter table, which represents a summary of results in all species represented in TriTrypDB, provides a bird's-eye view of all results and allows quick access to those results by simply clicking on the cells in the table. (**D**) Tabular representation of results [highlighted in yellow in the strategy in (A) and table in (C)—this table is interactive allowing the addition, deletion and reordering of columns]. Results of searches may be downloaded by clicking on the download results link (red circle).

Viewing a gene page can be achieved by either entering a specific gene ID in the ID search window in the interactive banner (Figure 1A), or the ID search query (Figure 1C) or by clicking on the gene ID in the results table of a search strategy (Figure 2D). The gene page (Figure 3A) contains all available information for a gene displayed on a single page, including synteny maps (Figure 3B), information on orthologs and paralogs (Figure 3C), EC (enzyme commission) numbers and genome ontology associations (Figure 3D), proteomics (Figure 3E), microarray and EST data (Figure 3F) and the actual sequence (amino acid and nucleotide) of the displayed gene (data not shown). In addition, links to User Comments (and access to the gene-specific comment form; green insert in Figure 3A) and linkouts to the gene record on GeneDB (blue insert in Figure 3A) are available through the gene page. Similarly, GeneDB provides links to appropriate records in TriTrypDB.

CURATION

We have implemented a synergistic approach to annotation, integrating the staff and expertise available as part of the GeneDB and EuPathDB projects with invaluable help from the broader scientific community. Annotators and curators at three sites [Seattle Biomedical Research Institute (USA), University of Georgia, Athens (USA) and the Wellcome Trust Sanger Institute (UK)] are all able to remotely curate, using Virtual Private Network connections to the Gene Builder interface of the Artemis annotation tool (17), which reads and writes directly to a Chado relational database at GeneDB (18). Curation currently focuses on preparing new sequence releases, updates to gene structure and function annotation and mutant phenotype annotations. This process is aided by providing members of the trypanosomatid research community the ability to directly add annotations to genes in TriTrypDB in the form of User Comments (green insert in Figure 3A). Comments made by scientific community members are forwarded to the annotators, in addition to immediately appearing on the gene record page in TriTrypDB. The comment form is designed to allow the user to input information into structured fields helpful to an annotator (in essence an expert opinion to guide the annotator), such as synonyms, experimentally-validated gene coordinates, gene product functional characterization, PubMed IDs, GenBank accession numbers and related genes (whereby the comment is replicated on related gene pages). We have found this to be a valuable forum for community input into TriTrypDB, of considerable use to curators working to improve gene annotations. Updates, committed to the GeneDB Chado database,



Download
Show All
Hide All

LinJ20_V3.0120
phosphoglycerate kinase B, cytosolic
View updated annotation at GeneDB

Overview
L. infantum protein coding gene on Lin.chr20 (chromosome 20) from **36634** to **37911**

Genomic Context Show [Data Sources]
Multiple Sequence Alignment Show
External Links Show
User Comments
none Add a comment on LinJ20_V3.0120 **Note: record updated** (View updated annotation at GeneDB)

External Links Hide

Database
GeneDB

Notes Show
Phenotype *none*
Orthologs and Paralogs within TriTrypDB Show [Data Sources]
EC Number Show [Data Sources]
GO Terms Show [Data Sources]

Protein Features Show [Data Sources]
Predicted Molecular Weight
45743 Da
Isoelectric Point
8.7
Similarities to Protein Data Bank (PDB) Chains Show [Data Sources]
Epitopes from IEDB *none* [Data Sources]

Add a comment to gene Tb927.1.710
Please add only scientific comments to be displayed on the gene page for Tb927.1.710. If you want to report a problem, use the support page. Your comments are appreciated. They will be forwarded to the Annotation Center for review and possibly included in future releases of the genome. If this is a **new gene**, please click here.
Part I: Comment
Headline *
Category ☐ Gene Model ☐ Name/Product ☐ Function ☐ Expression ☐ Sequence ☐ Phenotype
Comment *
Location Strand: ● Forward ○ Reverse
Genome Coordinates: 234176-235438 ▸help
Part II: Evidence for This Comment (Optional)
Upload File Select a file: Choose File no file selected
Brief Description: (4000 max characters)
Add Another File
PubMed ID(s) ▸help
Preview
Genbank Accession(s) ▸help
Part III: Other Genes to which you want to apply this comment (Optional)
Gene Identifiers ▸help
Add Comment

Expression

Differentiation Time Series: Promastigote to Amastigote Show [Data Sources]

Sequences

Please note that UTRs are not available for all gene models and may result in the RNA sequence (with introns removed) being identical to the CDS in those cases.
Predicted Protein Sequence Show
Predicted RNA/mRNA Sequence (introns spliced out) Show
Genomic Sequence (introns shown in lower case) Show
Coding Sequence Show
Genome Sequencing and Annotation
Sequence data for *Leishmania infantum* clone JPCM5 (MCAN/ES/98/LLM-877) were downloaded from GeneDB (sequence and annotated features).
Sequencing of *L. infantum* was conducted by The Sanger Institute pathogen sequencing unit.

Lin.chr20
30k
40k
50k
Annotated Genes (with UTRs when available)
LinJ20_V3.0080
LinJ20_V3.0090
LinJ20_V3.0100
LinJ20_V3.0110
LinJ20_V3.0120
LinJ20_V3.0130
LinJ20_V3.0140
LinJ20_V3.0150
LinJ20_V3.0160
L. major genomic sequences
L. major Syntenic Genes
L. braziliensis genomic sequences
1640k
1650k
1660k
L. braziliensis Syntenic Genes
T. brucei genomic sequences
340k
350k
360k
T. brucei Syntenic Genes
T. cruzi non-esmeraldo genomic sequences
140k
150k
160k
T. cruzi non-esmeraldo Syntenic Genes.
T. cruzi esmeraldo genomic sequences
130k
140k
150k
T. cruzi esmeraldo Syntenic Genes
Unified MS/MS Peptides (see <10k region for details)

Orthologs and Paralogs within TriTrypDB Hide

Gene	Species	Product	is syntenic	has comments
LbrM20_V2.4190	Leishmania braziliensis	PGKC, phosphoglycerate kinase C	yes	no
LinJ20_V3.0110	Leishmania infantum	phosphoglycerate kinase C, glycosomal	no	no
LmjF20.0110	Leishmania major	phosphoglycerate kinase B, cytosolic	yes	no
LmjF20.0100	Leishmania major	phosphoglycerate kinase C, glycosomal	yes	no
Tb927.1.700	Trypanosoma brucei TREU927	phosphoglycerate kinase	yes	no
Tb927.1.710	Trypanosoma brucei TREU927	phosphoglycerate kinase	yes	no
Tc00.1047053505999.100	Trypanosoma cruzi strain CL Brener	3-phosphoglycerate kinase, glycosomal	yes	no
Tc00.1047053511419.40	Trypanosoma cruzi strain CL Brener	phosphoglycerate kinase, putative	yes	no
Tc00.1047053511419.50	Trypanosoma cruzi strain CL Brener	phosphoglycerate kinase, putative	yes	no
Tc00.1047053505999.90	Trypanosoma cruzi strain CL Brener	phosphoglycerate kinase, putative	yes	no

EC Number Hide

EC Number	Source	EC Description
2.7.2.3	Annotation Center	Phosphoglycerate kinase

GO Terms Hide

Ontology	GO ID	GO Term Name	Source	Evidence Code	Is Not
Biological Process	GO:0006096	glycolysis	Interpro	IEA	
Molecular Function	GO:0004618	phosphoglycerate kinase activity	Interpro	IEA	

Differentiation Time Series: Promastigote to Amastigote Hide

Figure 3. Screen shot of a gene record page in TriTrypDB. (**A**) The gene page, with all data ‘hidden’; any available data type can be viewed by clicking on ‘show’. Display preferences (and prior queries) are saved for registered users. (**B**) Genomic context view, showing SynView (19) synteny map between organisms supported in TriTrypDB. (**C**) Orthologs and paralogs table. (**D**) Tables representing EC (enzyme commission) numbers and GO (Genome Ontology) associations. (**E**) Protein features, including mapped peptides from proteomics experiments, InterPro domain predictions, hydropathy plots and BlastP results. (**F**) Evidence of transcript expression, from microarray experiments.

set a flag on the relevant page in TriTrypDB, alerting users to updated annotations, a process mediated by web-services. These changes are propagated to TriTrypDB as part of the subsequent data update and release. To date, this collaborative effort has yielded modifications of 1159 gene records, including changes to the gene product name, gene structure, addition of untranslated regions and the addition of functional information and PubMed citations.

FUTURE DIRECTIONS

TriTrypDB will continue to expand both in functionality and data content over the coming years. Data types for which we anticipate storing and providing a query interface include expression data from proteomics experiments and transcriptome analysis (by microarrays and RNA-Seq); DNA-binding data (ChIP-chip and ChIP-seq); metabolomic data and metabolic pathway reconstructions; new genome sequences and annotation, including genome variation data; and re-assemblies of current genome sequences.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contribution of numerous members of the trypanosomatid research community, in the form of advice, suggestions and/or data—often made available to the community via TriTrypDB in advance to publication.

FUNDING

Grant from the Bill & Melinda Gates Foundation to develop a TriTrypDB component of the EuPathDB project (Grant Number 50097 to D.S.R., J.C.K., C.J.S. and P.J.M.); Wellcome Trust (Grant Number WT085822MA to M.C., D.F.S., M.B. and D.S.R.); Core funding from of the Wellcome Trust Sanger Institute by the Wellcome Trust (Grant Number WT085775/Z/08/Z). EuPathDB, which provides the infrastructure upon which TriTrypDB was constructed, is funded with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No.

HHSN266200400037C (to D.S.R., C.J.S. and J.C.K.). Funding for open access charge: Bill and Melinda Foundation (Grant Number 50096) and The Wellcome Trust (Grant Number WT085775).

Conflict of interest statement. None declared.

REFERENCES

1. Cox,F.E. (2004) History of sleeping sickness (African trypanosomiasis). *Infect. Dis. Clin. North Am.*, **18**, 231–245.
2. The World Health Organization. (2006) Human African trypanosomiasis (sleeping sickness): epidemiological update. *Wkly. Epidemiol. Rec.*, **81**, 71–80.
3. Simarro,P.P., Jannin,J. and Cattand,P. (2008) Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med.*, **5**, e55.
4. Hotez,P.J., Bottazzi,M.E., Franco-Paredes,C., Ault,S.K. and Periago,M.R. (2008) The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Negl. Trop. Dis.*, **2**, e300.
5. World Health Organization. Leishmaniasis: magnitude of the problem. World Health Organization, Geneva, 2009. http://www.who.int/leishmaniasis/burden/magnitude/burden_magnitude/en/index.html (July 2009, last date accessed).
6. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E., Haas,B. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*, **309**, 416–422.
7. El-Sayed,N.M., Myler,P.J., Bartholomeu,D.C., Nilsson,D., Aggarwal,G., Tran,A.N., Ghedin,E., Worthey,E.A., Delcher,A.L., Blandin,G. *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*, **309**, 409–415.
8. Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R. *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, **309**, 436–442.
9. Peacock,C.S., Seeger,K., Harris,D., Murphy,L., Ruiz,J.C., Quail,M.A., Peters,N., Adlem,E., Tivey,A., Aslett,M. *et al.* (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat. Genet.*, **39**, 839–847.
10. Hertz-Fowler,C., Peacock,C.S., Wood,V., Aslett,M., Kerhornou,A., Mooney,P., Tivey,A., Berriman,M., Hall,N., Rutherford,K. *et al.* (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.*, **32**, D339–D343.
11. Aurrecoechea,C., Heiges,M., Wang,H., Wang,Z., Fischer,S., Rhodes,P., Miller,J., Kraemer,E., Stoeckert,C.J. Jr, Roos,D.S. *et al.* (2007) ApiDB: integrated resources for the apicomplexan bioinformatics resource center. *Nucleic Acids Res.*, **35**, D427–D430.
12. Weatherly,D.B., Boehlke,C. and Tarleton,R.L. (2009) Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics*, **10**, 255.
13. Panigrahi,A.K., Ogata,Y., Zikova,A., Anupama,A., Dalley,R.A., Acestor,N., Myler,P.J. and Stuart,K.D. (2009) A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. *Proteomics*, **9**, 434–450.
14. Atwood,J.A. 3rd, Weatherly,D.B., Minning,T.A., Bundy,B., Cavola,C., Opperdoes,F.R., Orlando,R. and Tarleton,R.L. (2005) The *Trypanosoma cruzi* proteome. *Science*, **309**, 473–476.
15. Rosenzweig,D., Smith,D., Myler,P.J., Olafson,R.W. and Zilberstein,D. (2008) Post-translational modification of cellular proteins during *Leishmania donovani* differentiation. *Proteomics*, **8**, 1843–1850.
16. Chen,F., Mackey,A.J., Stoeckert,C.J. Jr. and Roos,D.S. (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.*, **34**, D363–D368.
17. Carver,T., Berriman,M., Tivey,A., Patel,C., Bohme,U., Barrell,B.G., Parkhill,J. and Rajandream,M.A. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics*, **24**, 2672–2676.
18. Mungall,C.J. and Emmert,D.B. (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.
19. Wang,H., Su,Y., Mackey,A.J., Kraemer,E.T. and Kissinger,J.C. (2006) SynView: a GBrowse-compatible approach to visualizing comparative genome data. *Bioinformatics*, **22**, 2308–2309.